Accepted Manuscript

Title: Prediction of Collision Cross Section and Retention Time for Broad Scope Screening in Gradient Reversed-Phase Liquid Chromatography-Ion Mobility-High Resolution Accurate Mass Spectrometry

Authors: Christian Brinch Mollerup, Marie Mardal, Petur Weihe Dalsgaard, Kristian Linnet, Leon Patrick Barron

PII: S0021-9673(18)30189-4
DOI: https://doi.org/10.1016/j.chroma.2018.02.025
Reference: CHROMA 359204

To appear in: *Journal of Chromatography A*

Received date: 15-12-2017
Revised date: 6-2-2018
Accepted date: 14-2-2018

Please cite this article as: Christian Brinch Mollerup, Marie Mardal, Petur Weihe Dalsgaard, Kristian Linnet, Leon Patrick Barron, Prediction of Collision Cross Section and Retention Time for Broad Scope Screening in Gradient Reversed-Phase Liquid Chromatography-Ion Mobility-High Resolution Accurate Mass Spectrometry, Journal of Chromatography A https://doi.org/10.1016/j.chroma.2018.02.025

Prediction of Collision Cross Section and Retention Time for Broad Scope Screening in Gradient Reversed-Phase Liquid Chromatography-Ion Mobility-High Resolution Accurate Mass Spectrometry

Christian Brinch Mollerup[1*], Marie Mardal[1], Petur Weihe Dalsgaard[1], Kristian Linnet[1], Leon Patrick Barron[2]

[1]Section of Forensic Chemistry, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's vej 11, 3. DK-2100, Denmark.

[2]Analytical & Environmental Sciences Division, Faculty of Life Sciences & Medicine, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, United Kingdom

*Corresponding author: Christian Brinch Mollerup
Section of Forensic Chemistry, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's vej 11, 3. DK-2100, Denmark. Phone: +4523436333, E-mail: christian.brinch.mollerup@sund.ku.dk

Highlights

- Retention time (RT) and collision cross section (CCS) prediction of small-molecule drugs
- Single and combined artificial neural network prediction models of RT and CCS
- Prediction errors evaluated with external validation set
- 91.9% within both 2 minutes RT error and 5% relative CCS (combined ANN model)

Abstract

Exact mass, retention time (RT), and collision cross section (CCS) are used as identification parameters in liquid chromatography coupled to ion mobility high resolution accurate mass spectrometry (LC-IM-HRMS). Targeted screening analyses are now more flexible and can be expanded for suspect and non-targeted screening. These allow for tentative identification of new compounds, and *in-silico* predicted reference values are used for improving confidence and filtering false-positive identifications. In this work, predictions of both RT and CCS values are performed with machine learning using artificial neural networks (ANNs). Prediction was based on molecular descriptors, 827 RTs, and 357 CCS values from pharmaceuticals, drugs of abuse, and their metabolites. ANN models for the prediction of RT or CCS separately were examined, and the potential to predict both from a single model was investigated for the first time. The optimized combined RT-CCS model was a four-layered multi-layer perceptron ANN, and the 95th prediction error percentiles were within 2 minutes RT error and 5% relative CCS error for the external validation set (n=36) and the full RT-CCS dataset (n=357). 88.6% (n=733) of predicted RTs were within 2 minutes error for the full dataset. Overall, when using 2 minutes RT error and 5% relative CCS error, 91.9% (n=328) of compounds were retained, while 99.4 % (n=355) were retained when using at least one of these thresholds. This combined prediction approach can therefore

be useful for rapid suspect/non-targeted screening involving HRMS, and will support current workflows.

Keywords: collision cross section prediction; retention time prediction; artificial neural networks

1. Introduction

Liquid chromatography coupled to high resolution accurate mass spectrometry (LC-HRMS) has enabled comprehensive toxicological screening of large numbers of trace contaminants in complex matrices such as biological samples and environmental matrices [1–6]. The addition of ion mobility spectrometry (IMS) has recently represented a significant increase in capability and allows for separation of ions in the gas-phase based on their mobility differences in an applied electric field [7,8]. Ions are then measured by their drift times through a tube containing a buffer gas. While drift times are system dependent, the average collision cross sections (CCS) between the ion and buffer gas can be derived when using constant operating procedures. The CCS of an ion is correlated to its size, shape, and charge. After calibration, the drift times observed from a travelling-wave IMS (TW-IMS) system can be used to determine CCS values [9]. CCS from TW-IMS have been shown to be matrix and system independent [10,11], and the use of LC-IMS-HRMS have been used to reduce the number of false-positive identifications and can replace other screening metrics for confirmatory analysis as a result (e.g. isotopic pattern match and fragment ions) [12]. The use of RT and CCS for confirmatory analyses means there also exists a lesser need for data-dependent fragmentation as the full-scan HRMS fragmentation can be filtered both on RT and drift time alignment [7]. This can then be applied to targeted, suspect, and non-targeted screening as required using the same dataset.

A common challenge, particularly in forensic screening, is keeping methods updated with relevant compounds. More than two new psychoactive substances enter the American and European drug market every week, on average [13,14]. Also, with the increase of long-distance travel for vacations and work, local populations can be exposed to pollutants and drugs not prescribed in their home countries. Suspect and non-targeted screening approaches have been utilized for identification of

compounds before acquisition of reference standards [4,15–17]. For this purpose, *in-silico* fragmentation matching [18,19] and prediction of retention time (RT) have been shown to reduce the list of potential compounds [1,20]. *In-silico* prediction of CCS and IMS drift times have utilized molecular modelling techniques [21–23]; however, models based on molecular descriptors have shown similar results while drastically reducing computing time [24–27], which corresponds to findings for prediction of the reduced ion mobility constants [28,29].

The aim of this work was to predict both RT and CCS with the use of artificial neural networks (ANNs), a machine learning technique that has been demonstrated for predicting analytical reference values, and has only very recently been utilized for prediction of either RT [1,30] or CCS [24] for use in screening. However, combination of these tools to understand their added value for preliminary suspect identifications has not yet been performed. A previously developed ANN model for RT prediction was trained and validated herein on a new, significantly larger dataset gathered under different LC conditions and in a different laboratory; ANN and linear regression models for prediction of CCS were trained and validated, and finally a combined model for prediction of both RT and CCS simultaneously was critically evaluated. This novel approach to *in silico* prediction of both RT and CCS alongside the use of HRMS data will markedly increase the speed and confidence in tentative identifications of potentially large numbers of new compounds.

2. Materials and methods

2.1. Chemicals

Reference standards of pharmaceuticals, drugs of abuse, and their metabolites were purchased from Lipomed GmbH (Bad Säckingen, Germany), Cerilliant (Round Rock, TX, USA), Toronto Research

Chemicals (Toronto, Canada), and SelleckChem (Houston, TX, USA). All reference standards were of ≥98% purity. Methanol, water, acetonitrile, propanol, and formic acid (LC-MS grade) were obtained from Fisher Scientific (Loughborough, UK). Leucine enkephalin was purchased from Sigma-Aldrich (Copenhagen, Denmark).

2.2. Instrumentation

Analyses were performed on two separate systems, an ultra-high performance liquid chromatography-time-of-flight mass spectrometer (UHPLC-TOF; System 1) and a UHPLC-TW-IMS-TOF (System 2). RT were obtained on System 1 with an ACQUITY UPLC I-Class coupled with a Xevo G2-S QTOF (Waters MS Technologies, Manchester, United Kingdom), and CCS values were obtained on System 2: an ACQUITY UPLC H-Class coupled with a VION IMS QTOF (Waters MS Technologies, Manchester, United Kingdom). LC separations on both systems were achieved using an Acquity UPLC HSS C_{18} column (150 mm × 2.1 mm, 1.8 μm), which was maintained at a constant temperature of 50 °C and a flow rate of 0.4 mL/min. Mobile Phase A consisted of 5 mM aqueous ammonium formate buffer adjusted to pH 3 with formic acid, and Mobile Phase B consisted of acetonitrile with 0.1% v/v formic acid. The gradient was 0 min to 0.5 min: 13% (B); from 0.5 min to 10 min: 13% to 50% (B); from 10 min to 10.75 min: 50% to 95% (B); from 10.75 min to 12.25 min: 95% (B); and from 12.25 min to 12.5 min: 95% to 13% (B); from 12.5 min to 15 min: 13% (B). The total run time was 15 min, and the injection volume was 3 μL. Ion mobility (System 2) was calibrated with a Major Mix IMS/Tof Calibration Kit from Waters, drift times were measured, and CCS values were calculated by the UNIFI software (Waters MS Technologies, Manchester, United Kingdom). Nitrogen (N_2) was used as drift gas in the TW-IMS of System 2.

With respect to mass spectrometry, both systems were used in positive electrospray ionization (Z-spray) mode with the following settings: nebulization gas 1000 L/h (System 1) and 800 L/h (System 2), with a desolvation temperature of 400 °C; cone gas flow 10 L/h (System 1) and 20 L/h (System 2); source temperature 150 °C; capillary voltage 800 V; cone voltage 25V; and argon as the collision gas. The low collision energy was set at 4 eV, and the high collision energy was ramped from 10 to 40 eV. The acquisition time was the entire run, with a scan time of 0.200 s. The minimum mass-to-charge (*m/z*) was 50 and the maximum was 950 (System 1) or 1000 (System 2). Mass calibration of System 1 was performed with 5 mM sodium formate solution in propanol: water (90:10, v/v), while System 2 was mass calibrated with the Major Mix IMS/Tof Calibration Kit from Waters. Lock mass was used with leucine enkephalin as a reference mass at *m/z* 556.2766 on both systems.

2.3. Reference values

In total, RTs for 869 compounds were determined from reference standards (Dataset I). Of these, the CCS of the proton adduct was determined for 364 compounds (Dataset II). For both datasets, compounds identified as multiple LC peaks were excluded. RTs were recorded on both systems, however, only RTs from System 1 was used for prediction. The differences in dataset sizes were primarily due to reference standards only being analyzed on System 1. Other factors were no observed protonation adducts, either due to high affinity for metal adducts or heavy in-source fragmentation.

2.4. Molecular descriptor generation

A total of 869 unique simplified molecular-input line-entry system (SMILES) strings were generated with ChemScript v16.0 from PerkinElmer (Waltham, MA, USA) from an in-house database of mol-files. Each SMILES string corresponded to a single compound and was used to generate a total of 105 molecular descriptors with Parameter Client freeware [31,32]. The selected descriptors were

constitutional descriptors, functional group counts, and molecular properties. Additional descriptors were generated for each compound: thirteen descriptors from ACD/Percepta (ACD/Labs, Toronto, Canada) and six descriptors from ChemScript v16.0 from PerkinElmer (Waltham, MA, USA). The full list of SMILES and corresponding descriptor values are available in Table A.1. Compounds for which the descriptor generation failed were excluded from the ANN modelling.

2.5. Descriptor selection and ANN optimization

ANN modelling was performed using Trajan Neural Networks v6.0. Prior to any evaluation, Dataset I & II were split into optimization and external validation sets with compounds chosen at random, in proportions 80:20 and 90:10, respectively. RT values for compounds exclusive to Dataset I were added as an external validation set in Models RT2 & RT-CCS (only regarding the RT prediction). The external validation set were used to reduce the risk of overfitting to the optimization set.

In total, four ANNs were trained and optimized. Single-output models for RT or CCS included Models RT1 & RT2, which were used with Dataset I & II, respectively, to predict RT, and Model CCS, which used Dataset II for predicting CCS. Model RT-CCS was a two-output model for predicting both CCS and RT simultaneously and used on Dataset II. The ANN was trained with backpropagation and conjugated gradient descent. Models RT1 & RT2 were based on 16 descriptors described by Barron and McEneff [33]. For Models CCS & RT-CCS, the number of descriptors was first reduced with feature selection and further reduced during the ANN optimization. Feature selection was performed by first removing duplicated descriptors and those with near-zero variance within the dataset. Near-zero variance was defined as having no variance or a higher than 95:5 ratio between most common and second-most-common descriptor values. Subsequently, for both models, feature selection with the Trajan software was repeated n=6 times each with forwards, backwards, and genetic selection. Only

descriptors selected in at least 16 of the 18 times in each case were retained as a priority given their linear correlation to CCS and/or RT. For the remaining descriptors, feature selection was repeated n=4 times with each method; only retaining those selected all twelve times. For the ANN optimization, the corresponding optimization set was split into the subsets: training, verification, and test set in a ratio of 70:15:15. The "Intelligent Problem Solver" of the Trajan software was used in four rounds per model. The ANN was trained with backpropagation and conjugated gradient descent. The outputs of the input nodes were scaled linear, and the hyperbolic activation function was used in the hidden nodes. In each round, choices were made based on prediction errors of the verification and test sets of the best network(s) from the previous round. The first round was used to select network type between radial basis functions, three/four-layer multi-layer perceptrons (MLPs), probabilistic neural networks, or generalized regression neural networks; with continuous resampling of training, verification, and (internal) test sets, and sub-selection of descriptors for Models CCS & RT-CCS. In the second round, only the optimal network type was used, where in the third round, network type, test set, and descriptors were fixed. In the fourth and final round, all but the network architecture (number of nodes in hidden layer) were locked. The final best network was then applied to the respective external validation set of the model. Further descriptor analysis was carried out with the built-in sensitivity analysis feature, which rates descriptors based on the deterioration in modeling performance when the descriptor is made unavailable.

2.6 Modelling CCS data by linear regression

Ordinary least squared regression of CCS as a function of molecular weight (MW) was performed for Dataset II (Model CCS-MW). The optimization set of Dataset II was split into training and test sets (70:30), and was resampled 250 times, each time making a model using linear regression of CCS as a

function of MW based on the training set. The model with the lowest 90th percentile prediction error for the full optimization set was selected and tested regarding the external validation set.

3. Results and Discussion

Thirty-one compounds were removed due to peak splitting of the early-eluting compounds (<1.1 min), and an additional eleven compounds were removed due to failed descriptor generation, 42 in total for Dataset I and seven for Dataset II. RT and CCS values, SMILES, and molecular descriptors are available in Table A.1 for included compounds. The molecular descriptors evaluated in this work were limited to the groups: constitutional descriptors, functional group counts, and molecular properties, comprising 124 molecular descriptors. However, the presented RT prediction of Models RT1 & RT2 was based upon previous work that had sampled a larger number of molecular descriptors [1,30,33].

3.1. Retention time modelling

All RT values were an average of at least four measurements obtained from mixtures of reference standards in solvents. RTs were not determined in spiked matrix samples since little to no influence of matrix on RTs (±0.02 min) have been observed with whole blood on System 1 [4]. In Table 1, key values of the prediction errors of all three RT models are given, and Figure 1 shows measured RT versus predicted RT. Predicted RT values for each compound is available in Table A.2. The optimized networks for Models RT1, RT2, & RT-CCS were four-layered MLPs. Like the previously reported contributions of molecular descriptors [33] to predictions of RT, logD, and atomic logP (AlogP; Ghose-Crippen octanol-water partition coefficient) were the most influential for all three models, with the number of carbons and number of oxygens being the third most influential for Models RT1 & RT2, respectively. The descriptors used in Model RT-CCS for prediction of RT were also used for CCS

prediction. MW and logD were the most influential, while number of unsubstituted benzene carbon and compound logP (ClogP) were the third and fourth most influential. Sensitivity analysis results and full list of descriptors are available in Table A.3.

Model RT1 showed improved prediction accuracy compared to Models RT2 & RT-CCS regarding the full external validation. This difference is most likely due to a larger optimization set and the difference in compounds available for optimization. Dataset I contained a larger variety of pharmaceutical compounds than Dataset II, and these were, therefore, not available in the optimization of Models RT2 & RT-CCS. This is shown by the difference in the prediction errors between the external validation sets when Dataset I was/was not included. This difference can be illustrated with principal component analysis based on the descriptors as shown in figure 2. The principal component analysis shows that models limited to Dataset II in the optimization have smaller applicability domain than those optimized on Dataset I. Model RT-CCS showed a small improvement over Model RT2, which can be explained by the descriptors being selected specifically for this dataset and, therefore, this analytical system. In general, this shows the need to retrain ANN prediction models when the experimental values are system dependent and the limitation that only compounds similar to the compounds of the optimization can be expected to fall within prediction error tolerances.

3.2. Collisional Cross-section modelling

The CCS of the protonated adduct was for all compounds in the range 128-250 $Å^2$. All CCS values were an average of at least four measurements obtained from mixtures of reference standards in solvents. Predicted CCS values are available in Table A.4. A summary of the performance of models for CCS prediction is presented in Table 2. The linear regression using MW provided a good prediction; however, the ANN models improved upon this, and a less than 5% relative prediction error

for 95% of the external validation set was achieved with Models CCS & RT-CCS. Model CCS used eight molecular descriptors: MW, parachor, number of bonds, number of hydrogens, and the Wiener index. All these describe the size of the molecule. The last three descriptors were Ghose-Viswanadhan-Wendoloski drug-like indices: Psychotic-80, Inflammat-50 and Infective-80 [34]. Figure 3 shows the measured versus the predicted CCS. In total, 66.4% (n=237) and 61.1% (n=218) were within 2% relative error for Models CCS & RT-CCS, respectively. Model CCS performed better than RT-CCS on the optimization set (which was comprised of the same compounds), while prediction accuracy was similar for the external validation set. This difference is likely due to compromises in the optimization of Model RT-CCS, since the selection of descriptors and networks was based on prediction of both output variables. However, sensitivity analysis of Model RT-CCS showed that it relied mainly on a smaller number of descriptors for predictions such as, MW, LogD, nCbH, and ClogP. Generation of models using a smaller set of such descriptors failed to yield better or even similar prediction accuracy and this was not considered further. Generation of the larger descriptor dataset did not add significantly to the time required for predictions, as several thousand descriptors could be calculated simultaneously in a matter of minutes. ANN predictions using an optimized model took seconds. Depending on the ANN software, the combined RT-CCS model is advantageous when both values are desired, since maintaining a single ANN and predicting new values is simplified.

Other studies have predicted RT [1] and CCS [24] with ANNs, albeit separately, and found they could add confidence to identifications in suspect and non-targeted screening. However, our CCS prediction errors are slightly lower than those of Bijlsma et al. [24], most likely due to their smaller number of CCS cases used for optimization (205 vs 321), and their larger CCS value range (132-307 $Å^2$ vs 128-250 $Å^2$). That said, it was via the combination of both tools that the most significant advancement in *in*

silico capability was observed. Overall, the simultaneous RT-CCS model developed here achieved prediction accuracy within 2 minutes for 88.6% (n=733) of predicted RT values, and 95.2% (n=340) of predicted CCS values were within 5% relative error. Impressively, the true-positive rate was 91.9% (n=328) and 99.4% (n=355), when the criteria for both or either of predicted values were within their error limits, respectively. These limits retain a reasonable number of the true-positive identifications and represent a significant improvement in *in silico* predictive capability for screening which has not been demonstrated previously to our knowledge. However, the correlation between MW and CCS indicates a lack of accuracy amongst isomers and isobars, which is relevant since the predictions are intended as orthogonal identification criterions to accurate mass measurements. The predicted CCS were not orthogonal to accurate mass measurements in this study, as illustrated by the isomers, dosulepin and pizotifen. These isomers had a 4.7% relative CCS difference, which was the largest difference of CCS values between isomers in Dataset II. Tentative identification based on the predicted CCS values was possible when ranking was based on the prediction error; however, in both cases, they were within a 3% relative CCS error limit. When grouping Dataset II by exact mass with a tolerance of 3 mDa, 26 groups contain 2-4 compounds, in total, 60 compounds. The true-positive rate for these compounds were 96.7% (n=58) when applying both filters. However, the false-positive rate was 84.1% (n=74), and in 23.3% (n=14) of identifications a potential false-positive identification was removed. In two of these 14 cases, the false-positive identification was avoided based on predicted CCS. While this number shows room for improvement, it should be remembered these numbers are mostly based on filtering of structurally related isomers. When these limits are applied to large suspect and non-targeted databases, the true-positive rate will stay the same; while the false-positive rate will likely decrease. Similarly, when suspect and non-targeted screening are applied to authentic samples, the predicted

error filters can filter matrix components, which without filtering would be false-positive identifications.

Conclusion

Prediction models for RT, CCS, and a combination thereof have been presented. ANN prediction can support tentative identifications in suspect and non-target screening; however, limited selectivity between isomers and isobars were observed. The presented models include a large range of small-molecule compounds of toxicological interest, and prediction accuracies were reliant on the number and diversity of compounds used in optimization and testing. Overall, the use of simple molecular descriptors, which can be generated for new compounds in a matter of minutes, allowed for fast prediction of RT and CCS, enabling application in large compound databases like those used in suspect and non-targeted screening.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank Maria Maansson and Carsten Birk, and their colleagues at Chr. Hansen Holding A/S, for time on their instrument.

Appendices

Appendix A: Single Excel file (.xlsx) with four sheets with text referenced tables.

References

[1] R. Bade, L. Bijlsma, T.H. Miller, L.P. Barron, J.V. Sancho, F. Hernández, Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis, Sci. Total Environ. 538 (2015) 934–941.

[2] S. Broecker, S. Herre, B. Wüst, J. Zweigenbaum, F. Pragst, Development and practical application of a library of CID accurate mass spectra of more than 2,500 toxic compounds for systematic toxicological analysis by LC-QTOF-MS with data-dependent acquisition., Anal. Bioanal. Chem. 400 (2011) 101–17.

[3] A.G. Helfer, J.A. Michely, A.A. Weber, M.R. Meyer, H.H. Maurer, Liquid chromatography-high resolution-tandem mass spectrometry using Orbitrap technology for comprehensive screening to detect drugs and their metabolites in blood plasma, Anal. Chim. Acta. 965 (2017) 83–95.

[4] C.B. Mollerup, P.W. Dalsgaard, M. Mardal, K. Linnet, Targeted and non-targeted drug screening in whole blood by UHPLC-TOF-MS with data-independent acquisition, Drug Test. Anal. 9 (2017) 1052–1061.

[5] A.J. Pedersen, P.W. Dalsgaard, A.J. Rode, B.S. Rasmussen, I.B. Müller, S.S. Johansen, K. Linnet, Screening for illicit and medicinal drugs in whole blood using fully automated SPE and ultra-high-performance liquid chromatography with TOF-MS with data-independent acquisition, J. Sep. Sci. 36 (2013) 2081–2089.

[6] A.T. Roemmelt, A.E. Steuer, T. Kraemer, Liquid Chromatography, In Combination with a Quadrupole Time-of-Flight Instrument, with Sequential Window Acquisition of All Theoretical Fragment-Ion Spectra Acquisition: Validated Quantification of 39 Antidepressants in Whole Blood As Part of a Simultane, Anal. Chem. 87 (2015) 9294–9301.

[7] A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer, Practical application of in silico fragmentation based residue screening with ion mobility high-resolution mass spectrometry, Rapid Commun. Mass Spectrom. 31 (2017) 1147–1157.

[8] S. Stephan, J. Hippler, T. Köhler, A.A. Deeb, T.C. Schmidt, O.J. Schmitz, Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC+LC-IM-qTOF-MS using a CCS database, Anal. Bioanal. Chem. 408 (2016) 6545–6555.

[9] K. Giles, J.L. Wildgoose, D.J. Langridge, I. Campuzano, A method for direct measurement of ion mobilities using a travelling wave ion guide, Int. J. Mass Spectrom. 298 (2010) 10–16.

[10] G. Paglia, J.P. Williams, L. Menikarachchi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B.O. Palsson, G. Astarita, Ion

mobility derived collision cross sections to support metabolomics applications., Anal. Chem. 86 (2014) 3985–93.

[11] J. Regueiro, N. Negreira, M.H.G. Berntssen, Ion mobility-derived collision cross section as an additional identification point for multi-residue screening of pesticides in fish feed, Anal. Chem. 88 (2016) 11169–11177.

[12] J. Regueiro, N. Negreira, R. Hannisdal, M.H.G. Berntssen, Targeted approach for qualitative screening of pesticides in salmon feed by liquid chromatography coupled to traveling-wave ion mobility/quadrupole time-of-flight mass spectrometry, Food Control. 78 (2017) 116–125.

[13] European Monitoring Centre for Drugs and Drug Addiction and Europol, EU Drug Markets Report: Strategic Overview, EMCDDA–Europol Joint publications, Publications Office of the European Union, Luxembourg, 2016.

[14] United Nations Office on Drugs and Crime, World Drug Report 2016, Vienna, Austria, 2016.

[15] M. Ibáñez, J. V. Sancho, L. Bijlsma, A.L.N. van Nuijs, A. Covaci, F. Hernández, Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances, TrAC Trends Anal. Chem. 57 (2014) 107–117.

[16] A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer, Semi-targeted residue screening in complex matrices with liquid chromatography coupled to high resolution mass spectrometry: current possibilities and limitations., Analyst. 136 (2011) 1898–909.

[17] A. Pelander, P. Decker, C. Baessmann, I. Ojanperä, Evaluation of a high resolving power time-of-flight mass spectrometer for drug analysis in terms of resolving power and acquisition rate., J. Am. Soc. Mass Spectrom. 22 (2011) 379–85.

[18] E. Tyrkkö, A. Pelander, I. Ojanperä, Differentiation of structural isomers in a target drug database by LC/Q-TOFMS using fragmentation prediction., Drug Test. Anal. 2 (2010) 259–70.

[19] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra., BMC Bioinformatics. 11 (2010) 148.

[20] E. Tyrkkö, A. Pelander, I. Ojanperä, Prediction of liquid chromatographic retention for differentiation of structural isomers., Anal. Chim. Acta. 720 (2012) 142–8.

[21] V. D'Atri, M. Porrini, F. Rosu, V. Gabelica, Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure, J. Mass Spectrom. 50 (2015) 711–726.

[22] C. Lapthorn, F.S. Pullen, B.Z. Chowdhry, P. Wright, G.L. Perkins, Y. Heredia, How useful is molecular modelling in combination with ion mobility mass spectrometry for "small molecule" ion mobility collision cross-sections?, Analyst. 140 (2015) 6814–6823.

[23] L.C. Menikarachchi, S. Cawley, D.W. Hill, L.M. Hall, L. Hall, S. Lai, J. Wilder, D.F. Grant, MolFind: A Software Package Enabling HPLC/MS-Based Identification of Unknown Chemical Structures, Anal. Chem. 84 (2012) 9388–9394.

[24] L. Bijlsma, R. Bade, A. Celma, L. Mullin, G. Cleland, S. Stead, F. Hernandez, J.V. Sancho, Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis, Anal. Chem. (2017) acs.analchem.7b00741.

[25] G.B. Gonzales, G. Smagghe, S. Coelus, D. Adriaenssens, K. De Winter, T. Desmet, K. Raes, J. Van Camp, Collision cross section prediction of deprotonated phenolics in a travelling-wave ion mobility spectrometer using molecular descriptors and chemometrics, Anal. Chim. Acta. 924 (2016) 68–76.

[26] Z. Zhou, X. Shen, J. Tu, Z.-J. Zhu, Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility — Mass Spectrometry, Anal. Chem. (2016) acs.analchem.6b03091.

[27] M.T. Soper-Hopper, A.S. Petrov, J.N. Howard, S.-S. Yu, J.G. Forsythe, M.A. Grover, F.M. Fernández, V.A. Online, A.S. Petrov, J.N. Howard, M.A. Grover, F.M. Ferna, J.G. Forsythe, Collision cross section predictions using 2-dimensional molecular descriptors, Chem. Commun. (2017).

[28] M.D. Wessel, P.C. Jurs, Prediction of reduced ion mobility constants from structural information using multiple linear regression analysis and computational neural networks, Anal. Chem. 66 (1994) 2480–2487.

[29] M.D. Wessel, J.M. Sutter, P.C. Jurs, Prediction of reduced ion mobility constants of organic compounds from molecular structure., Anal. Chem. 68 (1996) 4237–43.

[30] T.H. Miller, A. Musenga, D.A. Cowan, L.P. Barron, Prediction of Chromatographic Retention Time in High-Resolution Anti-Doping Screening Data Using Artificial Neural Networks, Anal. Chem. 85 (2013) 10330–10337.

[31] I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E. V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V. V. Prokopenko, Virtual Computational Chemistry Laboratory – Design and Description, J. Comput. Aided. Mol. Des. 19 (2005) 453–463.

[32] VCCLAB, Virtual Computational Chemistry Laboratory, (2005). http://www.vcclab.org.

[33] L.P. Barron, G.L. McEneff, Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods, Talanta. 147 (2016) 261–270.

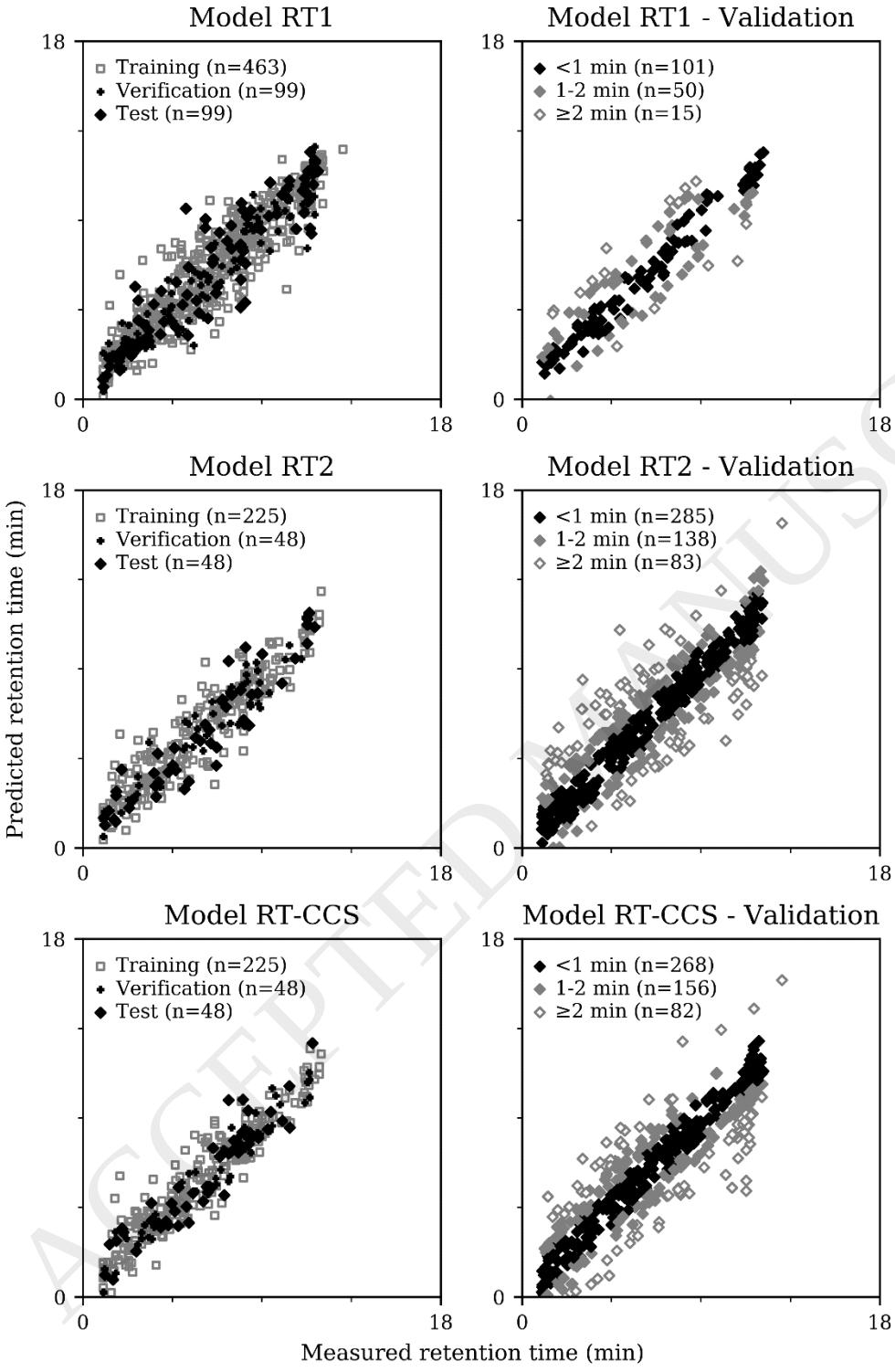
[34] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases, J. Comb. Chem. 1 (1999) 55–68.

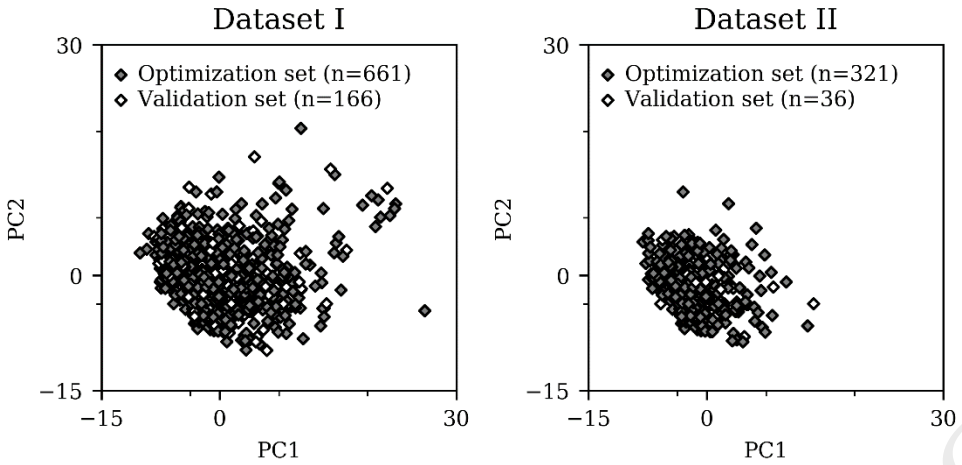
Figure Captions

Figure 1: Scatterplot of measured and predicted retention times for Models RT1, RT2, & RT-CCS, in the first, second and third row, respectively. The left column shows the optimization sets; training, verification, and test set. The right column shows the external validation, where points with less than 1 min error are black, absolute error between 1 min and 2 min are gray, and more than or equal to 2 min error are white.

Figure 2: Principal component analysis of the 124 descriptors, showing the difference in applicability domain.

Figure 3: Scatterplot of measured and predicted collision cross section for Models CCS-MW, CCS, & RT-CCS, in the first, second and third row, respectively. The left column shows the optimization set; training, verification, and test set. The right column shows the external validation, where points with less than 2% absolute relative error are black, less than 5% absolute relative error are grey, and more than or equal to 5% absolute relative error are white.





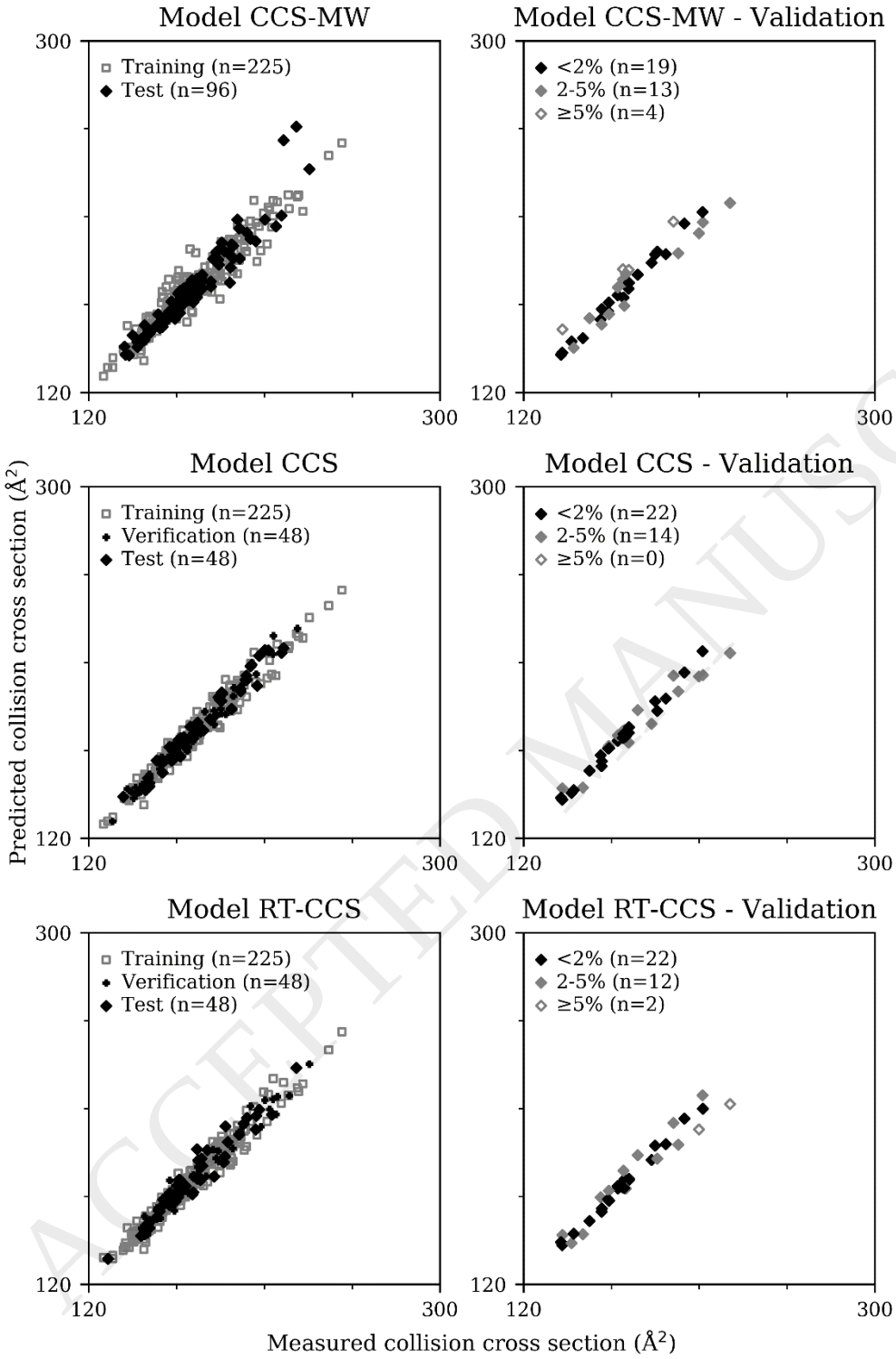


Table 1: Retention time prediction accuracy summary. Train, verification, and test sets are subsets of the optimization set.

Model	Best Network Architecture	Set	N	R^2	Slope (SE)	Intercept (SE)	Mean absolute error (min)	Median absolute error (min)	95th percentile (min)
RT1	MLP 16-14-5-1	Optimization	661	0.87	0.88 (0.01)	0.80 (0.10)	0.88	0.64	2.46
		Train	463	0.85	0.86 (0.02)	0.91 (0.12)	0.91	0.69	2.45
		Verification	99	0.92	0.89 (0.03)	0.57 (0.19)	0.71	0.50	1.86
		Test	99	0.87	0.91 (0.04)	0.60 (0.26)	0.88	0.53	2.91
		Validation	166	0.87	0.87 (0.03)	0.98 (0.18)	0.97	0.79	2.40
RT2	MLP 16-12-3-1	Optimization	321	0.87	0.88 (0.02)	0.70 (0.12)	0.80	0.57	2.14
		Train	225	0.86	0.86 (0.02)	0.82 (0.15)	0.81	0.59	2.18
		Verification	48	0.91	0.89 (0.04)	0.57 (0.29)	0.69	0.56	1.66

			Test	48	0.88	0.91 (0.05)	0.23 (0.33)	0.84	0.54	2.12
			Validation	36	0.85	0.85 (0.06)	0.37 (0.41)	1.03	0.85	2.21
			Validation *	506	0.81	0.86 (0.02)	1.03 (0.14)	1.13	0.82	3.09
RT-CCS	MLP	24-9-7-2	Optimization	321	0.90	0.90 (0.02)	0.57 (0.11)	0.73	0.60	1.89
			Train	225	0.89	0.89 (0.02)	0.65 (0.13)	0.74	0.59	1.98
			Verification	48	0.95	0.92 (0.03)	0.29 (0.24)	0.65	0.57	1.40
			Test	48	0.87	0.93 (0.05)	0.37 (0.35)	0.77	0.66	1.89
			Validation	36	0.87	0.86 (0.06)	0.59 (0.38)	0.88	0.75	1.93
			Validation *	506	0.82	0.86 (0.02)	0.73 (0.14)	1.14	0.92	2.85

*Additional compounds from dataset I included. MLP: multilayer perceptron; SE: standard error

Table 2: Collision Cross Section prediction accuracy overview. Train, verification, and test sets are subsets of the optimization set.

Model	Best Network Architecture	Set	N	R^2	Slope (SE)	Intercept (SE)	Mean absolute error (%)	Median absolute error (%)	95th percentile (%)
CCS-MW	n/a	Optimization	321	0.91	0.94 (0.02)	9.89 (2.95)	2.6	1.9	7.5
		Train	225	0.91	0.91 (0.02)	15.77 (3.37)	2.7	2.0	8.0
		Test	96	0.91	1.04 (0.03)	-7.01 (5.77)	2.2	1.6	4.8
		Validation	36	0.94	0.93 (0.04)	12.41 (7.32)	2.5	1.9	5.8
CCS	MLP 8-11-7-1	Optimization	321	0.97	0.97 (0.01)	4.56 (1.81)	1.7	1.3	4.2
		Train	225	0.96	0.96 (0.01)	7.38 (2.22)	1.7	1.2	4.5
		Verification	48	0.98	1.01 (0.02)	-0.86 (3.95)	1.5	1.3	3.4
		Test	48	0.97	1.01 (0.03)	-1.28 (4.82)	1.8	1.9	3.4
		Validation	36	0.96	0.93 (0.03)	12.31 (5.40)	1.8	1.7	4.0
RT-CCS	MLP 24-9-7-2	Optimization	321	0.96	0.96 (0.01)	7.59 (2.00)	1.9	1.5	5.0
		Train	225	0.96	0.95 (0.01)	7.90 (2.36)	1.9	1.5	5.1

Verification	48	0.97	0.95 (0.03)	9.21 (4.57)	1.7	1.3	4.1
Test	48	0.94	0.98 (0.04)	4.69 (6.26)	2.2	1.8	4.8
Validation	36	0.96	0.93 (0.03)	11.01 (5.89)	1.9	1.3	4.4

MLP: multilayer perceptron; SE: standard error